Z. Wahrscheinlichkeitstheorie verw. Gebiete
56, 113-126 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© Springer-Verlag 1981

Zero-Error Stationary Coding Over Stationary Channels

John C. Kieffer★

Department of Mathematics, University of Missouri-Rolla, Rolla, MO 65401, USA

Summary. For a type of stationary ergodic discrete-time finite-alphabet channel more general than the stationary totally ergodic $\bar{d}$-continuous channel of Gray, Ornstein and Dobrushin, it is shown that a stationary, ergodic source with entropy less than capacity can be transmitted over the channel with zero probability of error using stationary codes for encoding and decoding. This result generalizes the result of Gray et al. [3] that Bernoulli sources can be transmitted with zero error at rates below capacity over a totally ergodic $\bar{d}$-continuous channel.

1. Introduction

Let $[A, \mu]$ be an information source; the alphabet A of the source is always assumed finite, and μ, the distribution of the source, is a probability measure on the measurable space $(A^\infty, \mathscr{F}(A^\infty))$ consisting of A^∞, the set of all bilateral infinite sequences $x=(x_i)_{i=-\infty}^{\infty}$ from A, and $\mathscr{F}(A^\infty)$, the usual product σ-field of subsets of A^∞. We will assume that our source $[A, \mu]$ is both stationary (i.e., the shift T_A on A^∞ preserves μ), and ergodic (i.e., T_A-invariant sets have measure zero or one). We say that $[A, \mu]$ is aperiodic if $\mu(x)=0$ for every $x \in A^\infty$. We let $H(\mu)$ denote the entropy of the source $[A, \mu]$.

Let $[B, v, C]$ be a stationary channel, where the input alphabet B and the output alphabet C are finite, and $v=\{v_x: x \in A^\infty\}$ is a family of probability measures on C^∞ such that

(a) The map $x \to v_x(E)$ from $B^\infty \to [0,1]$ is $\mathscr{F}(B^\infty)$-measurable for each $E \in \mathscr{F}(C^\infty)$.

(b) $v_{T_B x}(T_C E)=v_x(E), x \in B^\infty, E \in \mathscr{F}(C^\infty)$.

We say the stationary and ergodic source $[A, \mu]$ is zero error transmissible over the stationary channel $[B, v, C]$ if there exist stationary codes $\varphi: A^\infty \to B^\infty$,

★ Research of author supported by NSF Grant MCS-78-21335 and the Joint Services Electronics Program under Contract N00014-79-C-0424.

0044-3719/81/0056/0113/$02.80

$\psi: C^\infty \to A^\infty$ and a Markov chain U, X, Y such that $U=\{U_i\}_{i=-\infty}^{\infty}$ is a process with state space A and distribution μ, X is the process with state space B such that $X=\varphi(U)$, Y is a process with state space C for which the distribution of Y conditioned on X is given by ν, and $U=\psi(Y)$ a.s. Intuitively speaking, if we encode the process U (which serves as a model for the information source $[A, \mu]$) into the process X, and then transmit X over the channel $[B, \nu, C]$, the process U can be recovered with probability one from the channel output process Y.

There is an equivalent way of formulating this. We say that the stationary sources $[A, \mu]$, $[B, \lambda]$ are isomorphic if there exist processes U, V which are stationary codings of each other such that U has distribution μ, and V has distribution λ. Following Gray et al. [3], given the stationary source $[B, \lambda]$ and the stationary channel $[B, \nu, C]$, we say $[B, \lambda]$ is ν-invulnerable if there are processes X, Y such that the distribution of X is λ, the distribution of Y conditioned on X is given by ν, and X is a stationary coding of Y. That is, $[B, \lambda]$ can be directly transmitted over the channel $[B, \nu, C]$ (without first encoding), and then recovered exactly from the channel output. It is not hard to see that $[A, \mu]$ is zero-error transmissible over $[B, \nu, C]$ if and only if there exists a ν-invulnerable source $[B, \lambda]$ isomorphic to $[A, \mu]$.

Shannon [9] showed that for a discrete memoryless channel there are two capacities C_0 and C (with $C_0 < C$ for most cases of interest) such that, if block encoders and decoders are used, C_0 is the maximum rate below which zero-error transmission is possible and C is the maximum rate below which transmission is possible with arbitrarily small (but possibly positive) probability of error. The number C is called the Shannon capacity and is equal to the supremum of the information rates of all stationary input-output measures for the channel; C_0 is called the zero-error capacity and has been calculated only for a few special cases considered by Shannon [9] and Lovasz [4] among others. One can see from an examination of Shannon's proof [9] that zero-error transmission using block coders is not possible at rates between C_0 and C because a block code has finite memory; that is, for some positive integer M, the output of the code at any time i is completely determined by looking at the sequence being coded at times $i-M$ through $i+M$. (See also [2], where some negative results are given on zero-error transmission using finite-memory sliding-block codes.) Gray et al. [3] showed that if infinite memory stationary encoders and decoders are used, zero-error transmission at any rate below C is possible provided the source being transmitted is a stationary coding of a memoryless source (i.e., a Bernoulli source), and the channel is totally ergodic and $\bar{d}$-continuous, a channel more general than the discrete memoryless channel.

In this paper, we show that for a type of channel more general than that considered by Gray et al. [3], zero-error transmission using stationary coders is possible for stationary, ergodic, aperiodic non-Bernoulli sources at all rates below C.

2. Principal Results

Let $\mathscr{P}_e(B)$ be the set of all probability measures on B^∞ for which the shift T_B is a measure-preserving, ergodic, aperiodic transformation on the probability space $(B^\infty, \mathscr{F}(B^\infty), \mu)$. On $\mathscr{P}_e(B)$ we place the unique metric topology for which convergence of a sequence of measures is weak convergence. Similarly, we place the topology of weak convergence on the set $\mathscr{P}(B, C)$ of all probability measures on $B^\infty \times C^\infty$. If λ is a probability measure on B^∞ and $[B, \nu, C]$ is a stationary channel, let $\lambda\nu$ be the probability measure on $B^\infty \times C^\infty$ such that

$$\lambda\nu(E \times F) = \int_E \nu_x(F)\, d\lambda(x),\ E \in \mathscr{F}(B^\infty),\ F \in \mathscr{F}(C^\infty).$$

Given a stationary channel $[B, \nu, C]$, let Φ_ν be the map $\lambda \to \lambda\nu$ from $\mathscr{P}_e(B) \to \mathscr{P}(B, C)$. If $\mu \in \mathscr{P}_e(\mathrm{B})$ and $[B, \nu, C]$ is a stationary channel, we say $[B, \nu, C]$ is ergodic at μ if the measure $\mu\nu$ is ergodic (with respect to the transformation $(x, y) \to (T_B x, T_C y)$ on $B^\infty \times C^\infty$).

In the following, if $\lambda \in \mathscr{P}_e(B)$ and $[B, \nu, C]$ is a stationary channel, $I(\lambda\nu)$ denotes the mutual information rate of the measure $\lambda\nu$ on $B^\infty \times C^\infty$.

For $k = 1, 2, \ldots$, the k-th order marginal distribution of $\lambda \in \mathscr{P}_e(B)$ is denoted by $\lambda^{(k)}$.

We now state the main result of the paper. (The proof is given in the last section.)

Theorem 1. *Let the stationary channel $[B, \nu, C]$ and $\tau \in \mathscr{P}_e(B)$ be given. Suppose there is some neighborhood $\mathscr{N}$ of τ in $\mathscr{P}_e(B)$ such that Φ_ν is continuous at every measure in $\mathscr{N}$ and $[B, \nu, C]$ is ergodic at every positive entropy measure in $\mathscr{N}$. Let $[A, \mu]$ be a stationary, ergodic, aperiodic source with $H(\mu) < I(\tau\nu)$. Then for any $k = 1, 2, \ldots$, and any $\delta > 0$, there exists a ν-invulnerable source $[B, \lambda]$ isomorphic to $[A, \mu]$ such that $\max_{x \in B^k} |\lambda^{(k)}(x) - \tau^{(k)}(x)| < \delta$. Thus, $[A, \mu]$ is zero-error transmissible over $[B, \nu, C]$.*

Definition. The Shannon capacity $C(\nu)$ of the stationary channel $[B, \nu, C]$ is the supremum of $I(\mu\nu)$ over all $\mu \in \mathscr{P}_e(B)$.

Corollary. *Let the stationary channel $[B, \nu, C]$ be ergodic at every positive entropy measure in $\mathscr{P}_e(B)$, and suppose Φ_ν is continuous at every measure in $\mathscr{P}_e(B)$. Let $[A, \mu]$ be any stationary ergodic aperiodic source. Then:*

(a) *$[A, \mu]$ is zero-error transmissible over $[B, \nu, C]$ if $H(\mu) < C(\nu)$, and*

(b) *$[A, \mu]$ is not zero-error transmissible if $H(\mu) > C(\nu)$.*

Part. Part (a) follows from Theorem 1. Part (b) was proved in [6].

As a special case, we get the results of Gray et al. [3], for, as shown in [6], the type of channel considered by these authors satisfies the hypotheses of the above corollary.

3. Synchronization Words

In order to construct the encoder for our source $[A, \mu]$, we will have to make sure that certain blocks of the encoder output are synchronization words; i.e., words which cannot be mistaken for cyclic shifts of themselves. In this section, for a given stationary and ergodic source we will show that there are synchronization words "typical" of the source.

Definition. Let $B^* = \bigcup_{n=1}^{\infty} B^n$, where B^n is the set of all n-tuples $(b_1, \ldots, b_n)$ from B. (B^* is thus the set of all finite-length words whose letters are from the alphabet B.) Define $\pi: B^* \to B^*$ to be the map

$$\pi(b_1, b_2, \ldots, b_n) = (b_2, \ldots, b_n, b_1).$$

If S is a set of integers and m is a positive integer we say S has minimal distance m if $|i-j| \geqq m$ for every $i, j \in S$, $i \neq j$. From now on, if G is a finite set, $|G|$ denotes the cardinality of G.

If A is a finite set, and $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$ are in A^n, let $d(x, x')$ denote the Hamming distance between these sequences:

$$n^{-1} |\{1 \leqq i \leqq n: x_i \neq x'_i\}|.$$

Lemma 1. *Let m be a positive integer. Let $\{X_i\}_{-\infty}^{\infty}$ be a stationary ergodic process with state space B such that if S is a set of integers with minimal distance m, then $\{X_i: i \in S\}$ are independent. Let $\lambda = \inf_{i \neq 0} Pr\{X_i \neq X_0\}$. Then, for any $\varepsilon > 0$,*

$$\lim_{n \to \infty} Pr[\min_{1 \leqq k \leqq n-1} d((X_1, \ldots, X_n), \pi^k(X_1, \ldots, X_n)) \leqq \lambda - \varepsilon] = 0.$$

Proof. See the proof of Lemma A2 of [8].

Definition. Let Z denote the set of integers. Let S be a subinterval (possibly infinite) of Z. Given a sequence of letters $x = \{x_i: i \in S\}$ from some alphabet, an integer k, and a positive integer r such that the interval $[k, k+r-1] = \{i \in Z: k \leqq i \leqq k+r-1\}$ is a subset of S, define $x_k^r = (x_k, \ldots, x_{k+r-1})$. If $k=0$, we write x^r for x_0^r. Similarly, if $X = \{X_i: i \in S\}$ is a sequence of random variables, define X_k^r and X^r.

Fix from now on $\tilde{X} = \{\tilde{X}_i: i \in Z\}$ to be the sequence of coordinate mappings from $B^\infty \to B$; $\bar{X} = \{\bar{X}_i\}$ to be the sequence of maps from $B^\infty \times C^\infty \to B$ such that $\bar{X}_i(x, y) = x_i$; and $\bar{Y} = \{\bar{Y}_i\}$ to be the sequence of maps from $B^\infty \times C^\infty \to C$ such that $\bar{Y}_i(x, y) = y_i$.

If μ is a T_B-stationary probability measure on B^∞ define $\lambda(\mu) = \inf_{i \neq 0} \mu[\tilde{X}_i \neq \tilde{X}_0]$. Also, if n, k are positive integers with $n > k$, and $\delta > 0$, we say $x \in B^n$ is (k, δ) typical of μ if for every $b \in B^k$, the distance between

$$n^{-1} |\{1 \leqq i \leqq n-k+1: x_i^k = b\}|$$

and $\mu^{(k)}(b)$ is less than δ.

Lemma 2. *Let $\mu \in \mathscr{P}_e(B)$. Then for any $\delta > 0$ and any positive integer m, there exists for n sufficiently large a subset W_n of B^n such that*

(a) $|W_n| \geqq 2^{n(H(\mu) - \delta)}$.

(b) *every element of W_n is (m, δ) typical of μ.*

(c) $\min_{1 \leqq k \leqq n-1} d(x, \pi^k x) \geqq \lambda(\mu) - \delta$ *for every $x \in W_n$.*

Proof. For each $n=1,2,\ldots$, let μ_n be the probability measure on B^∞ such that $\{\tilde{X}^n_{in}: i\in Z\}$ are independent under μ_n and each have distribution $\mu^{(n)}$. Let $\overline{\mu_n}$ be the probability measure $n^{-1}\sum_{i=0}^{n-1}\mu_n\cdot T_B^{ni}$ on B^∞. It is easily checked that $\overline{\mu_n}\in\mathscr{P}_e(B)$ and that if $S\subset Z$ has minimal distance n, $\{\tilde{X}_i: i\in S\}$ are independent under $\overline{\mu_n}$. Therefore, by Lemma 1, Lemma 2 holds for each $\overline{\mu_n}$. Since $\overline{\mu_n}\to\mu$ weakly, $H(\overline{\mu_n})\to H(\mu)$ and $\lambda(\overline{\mu_n})\to\lambda(\mu)$, Lemma 2 must also hold for μ.

Definition. If A is a finite set and $E\subset A^n$, we say E has minimal distance ε if $d(x,y)\geqq\varepsilon$ for every $x,y\in E$, $x\neq y$. If $x\in A^n$ and $1\leqq m\leqq n$ and $w\in A^m$, we say w is a m-subblock of x if $x_i^m=w$ for some i, $1\leqq i\leqq n-m+1$.

The following is proved by making a slight modification in the proof of Lemma A6 of [8].

Lemma 3. *Let A be a finite set and let $\varepsilon>0$. For $n=1,2,\ldots$, let $E_n\subset A^n$ be given with minimal distance 2ε. Suppose $\liminf_{n\to\infty} n^{-1}\log|E_n|>0$. Let k be a positive integer. For n sufficiently large, let $F_n\subset A^{kn}$ and a probability distribution p_n on F_n be given. Then for n sufficiently large, there exists $x_n\in E_n$ such that if*

$$F_n'=\{y\in F_n: d(x_n,c)\geqq\varepsilon \text{ for every } n\text{-subblock } c \text{ of } y\},$$

we have $p_n(F_n')\to 1$.

Definition. If A is a finite set and $0<\varepsilon\leqq\frac{1}{2}$, define

$$q_A(\varepsilon)=-\varepsilon\log\varepsilon-(1-\varepsilon)\log(1-\varepsilon)+2\varepsilon\log|A|,$$

where the logarithm is to base 2.

A subset of A^n of form $\{y\in A^n: d(x,y)<\delta\}$ for some $x\in A^n$ is called a Hamming ball of radius δ and center x.

Lemma 4. *Let $0<\varepsilon<\frac{1}{2}$. Then for n sufficiently large, every Hamming ball in A^n has cardinality no greater than $2^{nq_A(\varepsilon)}$, and every subset S of A^n has a subset S' of minimal distance ε such that $|S'|\geqq|S|2^{-nq_A(\varepsilon)}$.*

Proof. That the cardinality of every Hamming ball is bounded above as indicated, was shown in [5, p. 6]. Given $S\subset A^n$, find Hamming balls $B_1,\ldots,B_k$ of radius ε which cover S, such that the center of each ball is in S, and the center of B_j does not lie in $B_1\cup\ldots\cup B_{j-1}$, $1<j\leqq k$. Take S' to be the set of centers of the balls $\{B_i\}$.

The following is Lemma A3 of [8].

Lemma 5. *Let $\mu\in\mathscr{P}_e(B)$. Suppose $0<\varepsilon<\frac{1}{4}$ and $H(\mu)>q_B(\sqrt{\varepsilon})$. Then $\lambda(\mu)>\varepsilon$.*

Following is the synchronization lemma we will need later on to construct our source encoder.

Lemma 6. *Let $0<\varepsilon<\frac{1}{4}$ and let $\mu\in\mathscr{P}_e(B)$ satisfy $H(\mu)>q_B(\sqrt{\varepsilon})$. Let k,m be positive integers and let $\delta>0$. Then for n sufficiently large there exists $x\in B^n$ and $F\subset B^{kn}$ such that:*

(a) $\min_{1\leqq i\leqq n-1} d(x,\pi^i x)\geqq\varepsilon$.

(b) x is (m,δ) *typical of* μ.

(c) *every element of F is* (m, δ) *typical of* μ.
(d) *if* $y \in F$, *then* $d(x, c) \geqq \varepsilon$ *for every n-subblock c of y.*
(e) $|F| \geqq 2^{kn(H(\mu)-\delta)}$
(f) $\mu^{(kn)}(F) \to 1$ as $n \to \infty$.

Proof. Note that $\lambda(\mu) > \varepsilon$ by Lemma 5. Since $q_B(2\varepsilon) \leqq q_B(\sqrt{\varepsilon}) < H(\mu)$, we can choose $\eta > 0$ so small that $H(\mu) > \eta + q_B(2\varepsilon)$. By Lemmas 2 and 4, pick for n sufficiently large a set $E_n \subset B^n$ with minimal distance 2ε and cardinality at least $2^{n(H(\mu) - q_B(2\varepsilon) - \eta)}$ such that (a), (b) hold for every $x \in E_n$. For each n, choose $F_n \subset B^{kn}$ such that $\mu^{(nk)}(F_n) \to 1$ and every sequence in F_n is (m, δ) typical. Since $\liminf_{n \to \infty} n^{-1} \log |E_n| > 0$, by Lemma 3 we may find for n sufficiently large a $x \in E_n$ and $F \subset F_n$ such that (d) holds and $\mu^{(nk)}(F) \to 1$. Because (f) holds, (e) must hold for large n by the Shannon-McMillan Theorem.

4. Building Very Good Codes From Good Codes

Our method for proving Theorem 1 will work roughly this way. We first encode $[A, \mu]$ with a "good" code that produces small probability of error. Then, as a result of this section, we will be able to obtain a "very good" code by changing the original encoder a small amount, so that the new encoder produces a much smaller probability of error. In this way we construct a Cauchy sequence of better and better encoders such that the limit code is the zero-error code we seek.

Definition. If A is a finite set and n is a positive integer, we call $S \subset A^\infty$ a Rohlin n-set if the sets $S, T_A S, \ldots, T_A^{n-1} S$ are disjoint. We point out the following property of a Rohlin n-set $S \subset A^\infty$ for later use.

$$|\{p+1 \leqq i \leqq p+t \colon T_A^i u \in S\}| \leqq t n^{-1} + 1, u \in A^\infty, p \in Z, t = 1, 2, \ldots \tag{4.1}$$

We call a subset of A^∞ finite-dimensional (f.d.) if it is of form $\{u \in A^\infty \colon u_i^k \in S\}$ for some $i \in Z$, positive integer k and $S \subset A^k$. If G is another finite set, it should be clear what we mean by a f.d. subset of $A^\infty \times G^\infty$. (Make the obvious identification between $A^\infty \times G^\infty$ and $(A \times G)^\infty$.) We call a function from $A^\infty \to G$ f.d. if the pre-image of very element of G is a f.d. subset of A^∞. We call a map $\varphi \colon A^\infty \to G^\infty$ f.d. if for each $i \in Z$, the map $\varphi_i \colon A^\infty \to G$ is f.d., where

$$\varphi_i(u) = \varphi(u)_i, u \in A^\infty.$$

In the following, if S is a set, I_S denotes the indicator function of S. Also, d_w denotes the metric on $\mathscr{P}_e(B)$ yielding weak convergence such that

$$d_w(\mu, \lambda) = \sum_{n=1}^{\infty} 2^{-n} \sum_{x \in B^n} |\mu^{(n)}(x) - \lambda^{(n)}(x)|, \mu, \lambda \in \mathscr{P}_e(B).$$

The symbol T denotes the transformation $(x, y) \to (T_B x, T_C y)$ on $B^\infty \times C^\infty$.

Lemma 7. *Let $\mu\in\mathcal{P}_e(B)$ and let $[B,\nu,C]$ be a stationary channel which is ergodic at μ and for which Φ_ν is continuous at μ. Let $S\subset B^\infty\times C^\infty$ be f.d. and satisfy $\mu\nu(S)>\alpha>0$. Given τ $(0<\tau<1)$, there exist positive integers s, N and a $\delta>0$ such that if $n\geqq N$, if $W\subset B^n$ has every sequence in it (s,δ) typical of μ, if $\{\tilde{X}^{2n}\in W\times W\}$ is a Rohlin n-set, and if $\lambda\in\mathcal{P}_e(B)$ satisfies $n\lambda\{\tilde{X}^{2n}\in W\times W\}>1-\delta$, then*

(a) $\lambda\nu\left\{k^{-1}\sum_{i=0}^{k-1} I_{T^{-i}S}>\alpha\right\}>1-\tau, \qquad k\geqq N.$

(b) $d_w(\lambda,\mu)<\tau.$

Proof. Find α',τ' so that $\alpha'>0, 0<\tau'<1$, $\mu\nu(S)>\alpha'$, $\alpha'(1-\sqrt{\tau'})>\alpha$, $\sqrt{\tau'}<\tau$. Since $\mu\nu$ is ergodic, we may find J such that if G is the event $\left\{J^{-1}\sum_{j=0}^{J-1} I_{T^{-j}S}>\alpha'\right\}$, then $\mu\nu(G)>1-\tau'$. Since Φ_ν is continuous at μ, we may find $\beta>0$ such that if $\lambda\in\mathcal{P}_e(B)$ and $d_w(\lambda,\mu)<\beta$, then $\lambda\nu(G)>1-\tau'$. We can, and do, assume that $\beta<\tau$. Fix N,s,δ so that for $n\geqq N$:

(c) If $W\subset B^n$ has every sequence in it (s,δ) typical of μ, if $\{\tilde{X}^{2n}\in W\times W\}$ is a Rohlin n-set, and if $\lambda\in\mathcal{P}_e(B)$ satisfies $n\lambda(\tilde{X}^{2n}\in W\times W)>1-\delta$, then $d_w(\lambda,\mu)<\beta$.

(d) $$\sup_{B^\infty\times C^\infty}\left|(nJ)^{-1}\sum_{i=0}^{n-1}\sum_{j=i}^{i+J-1} I_{T^{-j}S}-n^{-1}\sum_{i=0}^{n-1} I_{T^{-i}S}\right|<\alpha'(1-\sqrt{\tau'})-\alpha.$$

Fix $k,n\geqq N$. Let $W\subset B^n$ and $\lambda\in\mathcal{P}_e(B)$ be given by (c). Then $E_{\lambda\nu}\left[k^{-1}\sum_{i=0}^{k-1} I_{T^{-i}G}\right]$ $=\lambda\nu(G)>1-\tau'$ and so by Chebyshev's inequality $k^{-1}\sum_{i=0}^{k-1} I_{T^{-i}G}>1-\sqrt{\tau'}$ with $\lambda\nu$-probability $>1-\sqrt{\tau'}$. This implies

$$\lambda\nu\left[(kJ)^{-1}\sum_{i=0}^{k-1}\sum_{j=1}^{i+J-1} I_{T^{-j}S}>\alpha'(1-\sqrt{\tau'})\right]>1-\sqrt{\tau'},$$

and hence by (d), (a) follows.

Definition. Given positive integers n,M and positive numbers ε, α such that $0<\varepsilon,\alpha<1$, a (n,M,ε,α) channel code for the stationary channel $[B,\nu,C]$ is a triple (W,G,g) where:

(a) $W\subset B^n, |W|=M$, and $\{\tilde{X}^{2n}\in W\times W\}$ is a Rohlin n-set.
(b) G is a f.d. Rohlin n-subset of C^∞.
(c) g is a f.d. map from $C^\infty\to W$.
(d) If $\lambda\in\mathcal{P}_e(B)$ and $n\lambda\{\tilde{X}^{2n}\in W\times W\}>1-\alpha$, then

$$n\lambda\nu[\bar{X}^{2n}\in W\times W, \bar{Y}\in G, \bar{X}^n=g(\bar{Y})]>1-\varepsilon.$$

The following lemma will allow us to construct a very good channel code from a good channel code.

Lemma 8. *Let $\mu\in\mathcal{P}_e(B)$, let the stationary channel $[B,\nu,C]$ be ergodic at μ, and let Φ_ν be continuous at μ. Let m,N be positive integers and let $F\subset B^{2m+1}$ be such that $\{\tilde{X}^{2m+1}\in F\}$ is a Rohlin N-set. Let $G\subset C^\infty$ be a f.d. Rohlin N-set, and* *let* $g: C^\infty \to B^N$ *be a f.d. function such that*

$$N\mu\nu[\bar{X}_{-m}^{2m+1}\in F, \bar{Y}\in G, \bar{X}^N = g(\bar{Y})] > 1-\delta,$$

where $\delta>0$ *is so small that* $\delta<(16)^{-4}$, $q_B(8\delta^{\frac{1}{4}})<H(\mu)$. *Then, given* $\tau(0<\tau<1)$, *there exists* $\alpha(0<\alpha<1)$ *such that for* N' *sufficiently large there exists a* (N', M, τ, α) *channel code* (W', G', g') *and a subset* Q *of* $B^{N'}$ *containing* W' *such that*

(a) $M \geqq 2^{N'(H(\mu)-q_B(8\delta^{\frac{1}{4}}))}$

(b) $\mu^{(N')}(Q)\to 1$, *and if* $x\in Q$ *there exists* $w\in W'$ *such that for more than* $N^{-1}N'(1-8\delta^{\frac{1}{4}})$ *of the integers* $i\in[m+1, N'-N-m]$ *one has* $x_{i-m}^{2m+1}, w_{i-m}^{2m+1}\in F$ *and* $w_i^N = x_i^N$.

(c) *If* $\lambda\in\mathscr{P}_e(B)$ *and* $N'\lambda\{\tilde{X}^{2N'}\in W'\times W'\}>1-\alpha$, *then* $d_w(\lambda,\mu)<\tau$.

Proof. To ease the notation, if $r,s\in Z$ and $r\leqq s$ and $x\in B^\infty$, $y\in C^\infty$, let $I_r^s(x,y)$ be the set of all $i\in[r,s]$ such that $x_{i-m}^{2m+1}\in F$, $T_C^i y\in G$, $g(T_C^i y)=x_i^N$. By Lemma 7, we may choose a positive $\alpha<\tau/2$, and positive integers s, J such that for every $n\geqq J$, if $W'\subset B^n$ has every sequence in it (s,α) typical of μ, if $\{\tilde{X}^{2n}\in W'\times W'\}$ is a Rohlin n-set, and $\lambda\in\mathscr{P}_e(B)$ satisfies $n\lambda\{\tilde{X}^{2n}\in W'\times W'\}>1-\alpha$, then

(d) $\lambda\nu[|I_0^{j-N}(\bar{X},\bar{Y})|\geqq N^{-1}j(1-\delta)]>1-\tau/2, j\geqq J$.

(e) $d_w(\lambda,\mu)<\tau$.

Define $\varepsilon=\sqrt{24\delta}$. Let k be the greatest integer in ε^{-1}. Since $q_B(\sqrt{\varepsilon})<H(\mu)$, by Lemma 6 for n sufficiently large there is $\tilde{x}\in B^n$ and $D\subset B^{kn}$ such that $\mu^{(kn)}(D)\to 1$ and:

(f) $\tilde{x}y\tilde{x}$ is (s,α) typical of μ, $y\in D$.

(g) $|D|\geqq 2^{kn(H(\mu)-\varepsilon)}$.

(h) $\min\limits_{1\leqq i\leqq n-1} d(\tilde{x},\pi^i\tilde{x})\geqq\varepsilon$.

(i) If $y\in D$, $d(\tilde{x},c)\geqq\varepsilon$ for every n-subblock c of y.

(j) $|\{m+1\leqq i\leqq kn-N-m\colon y_{i-m}^{2m+1}\in F\}|\geqq N^{-1}kn(1-\delta)$, $y\in D$.

Fix an arbitrary $n\geqq J$ for which (f)–(j) hold and

(k) $2\leqq n\delta N^{-1}$ and $2mn^{-1}<\delta$.

(l) Any Hamming ball in B^{nk} of radius 5ε has cardinality $\leqq 2^{nkq_B(5\varepsilon)}$.

Let σ be the symmetric reflexive relation on D such that

(m) $x\sigma y$ if and only if there are more than $N^{-1}kn(1-5\varepsilon)$ integers $i\in[m+1, kn-N-m]$ for which x_{i-m}^{2m+1}, $y_{i-m}^{2m+1}\in F$ and $x_i^N=y_i^N$.

Pick $D\subset D$ so that

(n) every $x\in D$ is σ-related to some $w\in D'$.

(o) if $w_1, w_2\in D'$ and $w_1\neq w_2$ then $w_1 \not\sigma w_2$.

Set $N'=2n+kn$, $W'=\{\tilde{x}y\tilde{w}\colon y\in D'\}$. Because of (*h*), (*i*), $\{\tilde{X}^{2N'}\in W'\times W'\}$ is a Rohlin N'-set. If $x\sigma y$, then x is in the Hamming ball of radius 5ε centered at y. Hence by (g), (l),

$$\begin{aligned}(N')^{-1}\log|W'| &\geqq (2n+kn)^{-1}kn(H(\mu)-q_B(6\varepsilon))\\ &\geqq (1-4\varepsilon)(H(\mu)-q_B(6\varepsilon))\geqq H(\mu)-q_B(8\varepsilon),\end{aligned}$$

from which (a) follows. Setting $Q=B^n\times D\times B^n$, then $Q\supset W'$, $\mu^{(N')}(Q)\to 1$, and since $N^{-1}kn(1-5\varepsilon)>N^{-1}N'(1-9\varepsilon)$, (b) follows from (n).

Let G' be the set of all $y \in C^\infty$ such that there exists $x \in B^\infty$ for which $x^{2N'} \in W' \times W'$ and

$$|I_0^{2N'-N}(x,y)| \geqq 2N^{-1}N'(1-2\delta).$$

Let $\bar{g}: C^\infty \to B^\infty$ be a f.d. stationary code such that $y \in G$ implies $\bar{g}(y)^N = g(y)$. If G' is not a Rohlin N'-set there exist $y \in C^\infty$ and $x, \bar{x} \in B^\infty$ and integers i,j with $i+1 \leqq j \leqq i+N'-1$ such that $x_i^{2N'}$, $x_j^{2N'} \in W' \times W'$ and

$$d(x_i^{2N'}, \bar{g}(y)_i^{2N'}) < 2\delta, d(\bar{x}_j^{2N'}, \bar{g}(y)_j^{2N'}) < 2\delta.$$

Because of (h), (i), there an integer r such that

$$[r, r+n-1] \subset [j, j+N'-1] \cap [i, i+2N'-1] \quad \text{and} \quad d(x_r^n, \bar{x}_r^n) \geqq \varepsilon.$$

Now $d(x_r^n, \bar{g}(y)_r^n) \leqq 4N'\delta n^{-1} < 12\varepsilon^{-1}\delta$.

Similarly, $d(\bar{x}_r^n, \bar{g}(y)_r^n) < 12\varepsilon^{-1}\delta$, and so $d(x_r^n, \bar{x}_r^n) < 24\varepsilon^{-1}\delta = \varepsilon$, a contradiction. We conclude from this that G' must be a Rohlin N'-set. Suppose $y \in C^\infty$, $x, \bar{x} \in B^\infty$, $x^{2N'}$, $\bar{x}^{2N'} \in W' \times W'$, and

$$|I_0^{2N'-N}(x,y)|, |I_0^{2N'-N}(\bar{x},y)| \geqq 2N^{-1}N'(1-2\delta).$$

Now by (4.1) and the fact (from (k)) that $N^{-1}N'+1 \leqq N^{-1}N'(1+\delta)$, we have

$$|I_0^{2N'-N}(x,y) \cup I_0^{2N'-N}(\bar{x},y)| \leqq 2N^{-1}N'(1+\delta),$$

whence

$$|I_0^{2N'-N}(x,y) \cap I_0^{2N'-N}(\bar{x},y)| \geqq 2N^{-1}N'(1-5\delta).$$

Applying (4.1) again and (k),

$$\begin{aligned}&|I_{n+m}^{(k+1)n-N-m-1}(x,y) \cap I_{n+m}^{(k+1)n-N-m-1}(\bar{x},y)| \\ &\quad \geqq 2N^{-1}N'(1-5\delta) - 2 - N^{-1}(N'+2n+2m+N) > N^{-1}N'(1-5\varepsilon),\end{aligned}$$

which implies $x_n^{kn} \sigma \bar{x}_n^{kn}$, and then $x^{N'} = \bar{x}^{N'}$ by (o). Thus there exists a map $g': C^\infty \to W'$ such that if $y \in C^\infty$ and $x \in B^\infty$ and $x^{2N'} \in W' \times W'$ and $|I_0^{2N'-N}(x,y)| \geqq 2N^{-1}N'(1-2\delta)$, then $g'(y) = x^{N'}$. Suppose $\lambda \in \mathscr{P}_e(B)$ and $N'\lambda(\tilde{X}^{2N'} \in W' \times W') > 1-\alpha$. Then (c) follows from (f). Also,

$$\begin{aligned}&N'\lambda v[\bar{X}^{2N'} \in W' \times W', \bar{Y} \in G', \bar{X}^{N'} \\ &= g'(\bar{Y})] = \lambda v\left[\bigcup_{i=0}^{N'-1} \{\bar{X}_i^{2N'} \in W' \times W', T_C^i \bar{Y} \in G', \bar{X}_i^{N'} = g'(T_C^i \bar{Y})\}\right] \\ &\geqq \lambda v\left[\bigcup_{i=0}^{N'-1} \{\bar{X}_i^{2N'} \in W' \times W', |I_i^{2N'-N+i}(\bar{X}, \bar{Y})| \geqq 2N^{-1}N'(1-2\delta)\right] \\ &\geqq \lambda v\left[\bigcup_{i=0}^{N'-1} \{\bar{X}_i^{2N'} \in W' \times W', |I_0^{3N'-N}(\bar{X}, \bar{Y})| \geqq 3N^{-1}N'(1-\delta)\right] \\ &\geqq N'\lambda(\tilde{X}^{2N'} \in W' \times W') + \lambda v[|I_0^{3N'-N}(\bar{X}, \bar{Y})| \geqq 3N^{-1}N'(1-\delta)] - 1 \\ &> (1-\tau/2) + (1-\tau/2) - 1 = 1-\tau.\end{aligned}$$

Hence, (W', G', g') is a $(N', |W'|, \tau, \alpha)$ channel code.

Notation. In the following if X, U are jointly stationary finite state processes $H(X|U)$ denotes the conditional entropy rate of X given U, and $H(X)$ denotes the entropy rate of X.

Lemma 9. *Let m, N be positive integers with $m > N$. Let A, B be finite sets. Let ε, δ be numbers such that $0<\varepsilon<1$, $0<\delta<1/3$. Given $F \subset A^{2m+1}$, $S_1 \subset S_2 \subset A^N$, $W \subset B^N$, $\varphi: S_1 \to W$, $\psi: W \to S_2$ and a jointly ergodic pair of processes U, X with respective state spaces A, B such that:*

(a) $\{u \in A^\infty : u_{-m}^{2m+1} \in F\}$, $\{x \in B^\infty : x^{2N} \in W \times W\}$ *are Rohlin N-sets.*

(b) *With probability* 1, $U_{-m}^{2m+1} \in F$, $U^N \in S_2$ *implies* $X^N \in W$, $\psi(X^N) = U^N$; $U_{-m}^{2m+1} \in F$, $U^N \in S_1$ *implies* $\varphi(U^N) = X^N$.

(c) $N Pr[U_{-m}^{2m+1} \in F, U_{N-m}^{2m+1} \in F, U^{2N} \in S_2 \times S_2] > 1 - \delta$;

$$N Pr[U_{-m}^{2m+1} \in F, U_{N-m}^{2m+1} \in F, U^{2N} \in S_1 \times S_1] > 1 - \varepsilon.$$

(d) *For N' sufficiently large there are subsets W', Q of $B^{N'}$ such that $W' \subset Q$, $Pr[X^{N'} \in Q] \to 1$, and for every $x \in Q$ there exists $w \in W'$ for which more than $N^{-1} N'(1-\delta)$ integers $i \in [1, N'-2N+1]$ satisfy $x_i^{2N}, w_i^{2N} \in W \times W$, $w_i^N = x_i^N$.*

(e) $H(X|U) > q_B(\delta) + q_A(3\delta)$.

Then for N' sufficiently large there is $V \subset A^{N'}$ such that $Pr[U^{N'} \in V] \to 1$ and a one-to-one map $\varphi': V \to W'$ with the property that:

For every $u \in V$, there are at least $N^{-1} N'(1 - 7\delta - \varepsilon)$ integers $i \in [m+1, N'-N-m]$ such that

$$u_{i-m}^{2m+1} \in F,\ u_{i+N-m}^{2m+1} \in F,\ u_i^{2N} \in S_1 \times S_1, \tag{4.2}$$

and $\varphi'(u)_i^N = \varphi(u_i^N)$.

Proof. Let $G \subset A^{N'} \times B^{N'}$ be the set of all (u, x) such that:

(f) $|\{m+1 \leqq i \leqq N'-N-m : u_{i-m}^{2m+1} \in F, u_{i+N-m}^{2m+1} \in F, u_i^{2N} \in S_2 \times S_2\}| \geqq N^{-1} N'(1-\delta)$

(g) $|\{m+1 \leqq i \leqq N'-N-m : u_{i-m}^{2m+1} \in F, u_{i+N-m}^{2m+1} \in F, u_i^{2N} \in S_1 \times S_1\}| \geqq N^{-1} N'(1-\varepsilon)$.

(h) For each $i \in [m+1, N'-N+1]$, $u_{i-m}^{2m+1} \in F$ and $u_i^N \in S_2$ imply $x_i^N \in W$,

$$\psi(x_i^N) = u_i^N;\ u_{i-m}^{2m+1} \in F \quad \text{and} \quad u_i^N \in S_1 \quad \text{imply } \varphi(u_i^N) = x_i^N.$$

(i) $x \in Q$ and $Pr[X^{N'} = x \mid U^{N'} = u] \leqq 2^{-N'(H(X|U) - \sigma)}$,

where $\sigma > 0$ is so small that $H(X|U) > q_B(\delta) + q_A(3\delta) + \sigma$. Let $V \subset A^{N'}$ be the set $\{u \in A^{N'} : Pr[X^{N'} \in G_u | U^{N'} = u] \geqq \frac{1}{2}\}$, where G_u denotes the section of G at u. Then, $Pr[U^{N'} \in V] \to 1$ as $N' \to \infty$. Fix N' so large that

(j) $2m + 1 + N < \delta N'$

(k) The Hamming balls in $A^{N'}$ of radius 3δ have no more than $2^{N' q_A(3\delta)}$ elements, and the Hamming balls in $B^{N'}$ of radius δ have no more than $2^{N' q_B(\delta)}$ elements.

$$r = [2^{N'(H(X|U) - \sigma - q_B(\delta))}/2] \geqq 2^{N' q_A(3\delta)} \geqq 1, \tag{1}$$

where if x is a real number, $[x]$ denotes the greatest integer in x. For each $u \in V$, let $A(u)$ be the set of all $\bar{u} \in V$ such that for more than $N^{-1}N'(1-3\delta)$ integers $i \in [m+1, N'-N-m]$ one has u_{i-m}^{2m+1}, $\bar{u}_{i-m}^{2m+1}$, u_{i+N-m}^{2m+1}, $\bar{u}_{i+N-m}^{2m+1} \in F$, and $u_i^{2N} \in S_2 \times S_2$ and $u_i^{2N} = \bar{u}_i^{2N}$. Pick $V' \subset V$ such that the sets $\{A(u): u \in V'\}$ cover V, and if $u_1, u_2 \in V'$ and $u_1 \neq u_2$, then $u_1 \notin A(u_2)$. For each $u \in V$ let $J(u)$ be the set of integers given in (f). Suppose $u_1, u_2 \in V'$ and $x \in G_{u_1} \cap G_{u_2}$. Now $J(u_1) \cup J(u_2)$ is contained in the set $\{m+1 \leqq i \leqq N'-N-m: x_i^{2N} \in W \times W\}$, whose cardinality is less than $N^{-1}N'(1+\delta)$ by (4.1) and (j). Hence, $|J(u_1) \cap J(u_2)| > N^{-1}N'(1-3\delta)$, which implies $u_1 \in A(u_2)$. Consequently, $G_{u_1} \cap G_{u_2} = \emptyset$ if $u_1, u_2 \in V'$ and $u_1 \neq u_2$. For each $x \in B^{N'}$, let $B(x)$ be the set of all $\bar{x} \in B^{N'}$ such that for more than $N^{-1}N'(1-\delta)$ integers $i \in [1, N'-2N+1]$ one has $x_i^{2N}, \bar{x}_i^{2N} \in W \times W$, and $x_i^N = \bar{x}_i^N$. Let $\tilde{V} \subset V'$. Let $G(\tilde{V}) = \bigcup\{G_u: u \in \tilde{V}\}$. Let $\{w_1, \ldots, w_t\} = W' \cap [\bigcup\{B(x): x \in G(\tilde{V})\}]$. If $x \in G(\tilde{V})$ then $x \in Q$ and so $w \in B(x)$ for some $w \in W'$ by (d). Thus $w \in \{w_1, \ldots, w_t\}$ and so $G(\tilde{V}) \subset \bigcup_{i=1}^{t} B(w_i)$. Now each $B(x)$ is contained in a Hamming ball of radius δ and therefore has cardinality at most $2^{N' q_B(\delta)}$ by (k). Also, each G_u has cardinality at least $2^{N'(H(X|U)-\sigma)}/2$. Therefore, $t \geqq |\tilde{V}| r$. By a marriage lemma [10, Lemma 2], there exist disjoint subsets $\{W_u': u \in V'\}$ of W', each having cardinality r, such that if $u \in V'$ and $w \in W_u'$, then there exists $x \in G_u$ with $w \in B(x)$. There are disjoint sets $\{A'(u): u \in V'\}$ which partition V such that $A'(u) \subset A(u)$, $u \in V'$. By (k), $|A'(u)| \leqq 2^{N' q_A(3\delta)}$ and so by (1) there is a one-to-one $\varphi': V \to W'$ such that $\varphi'(u) = w$ implies there exists $\bar{u} \in V'$ and $x \in B^{N'}$ for which $w \in B(x)$, $x \in G_{\bar{u}}$, and $u \in A(\bar{u})$. With $u, \bar{u}, w, x$ as just described, let

$$\begin{aligned}
J_1 &= \{m+1 \leqq i \leqq N'-N-m: x_i^{2N}, w_i^{2N} \in W \times W, w_i^N = x_i^N\} \\
J_2 &= \{m+1 \leqq i \leqq N'-N-m: \bar{u}_{i-m}^{2m+1}, \bar{u}_{i+N-m}^{2m+1} \in F, \bar{u}_i^{2N} \in S_1 \times S_1\} \\
J_3 &= \{m+1 \leqq i \leqq N'-N-m: \\
&\quad u_{i-m}^{2m+1}, \bar{u}_{i-m}^{2m+1}, u_{i+N-m}^{2m+1}, \bar{u}_{i+N-m}^{2m+1} \in F, u_i^{2N} = \bar{u}_i^{2N}, u_i^{2N} \in S_2 \times S_2\}.
\end{aligned}$$

Then, $|J_1| \geqq N^{-1}N'(1-2\delta)$, $|J_2| \geqq N^{-1}N'(1-\varepsilon)$, $|J_3| \geqq N^{-1}N'(1-3\delta)$. Note that if $i \in J_1 \cup J_2 \cup J_3$, then $x_i^{2N} \in W \times W$ and so by (4.1), $|J_1 \cup J_2 \cup J_3| \leqq N^{-1}N'(1+\delta)$. Thus,

$$|J_1 \cap J_2 \cap J_3| \geqq |J_1| + |J_2| + |J_3| - 2|J_1 \cup J_2 \cup J_3| \geqq N^{-1}N'(1-7\delta-\varepsilon).$$

Note that $i \in J_1 \cap J_2 \cap J_3$ implies that $u_{i-m}^{2m+1}, u_{i+N-m}^{2m+1} \in F$, and $u_i^{2N} \in S_1 \times S_1$, and $\varphi'(u)_i^N = \varphi(u_i^N)$, and so (4.2) follows.

Lemma 10. *Let $[A, \mu]$ be a stationary, ergodic, aperiodic source with entropy H. For N sufficiently large suppose we have sets $V \subset A^N$, $W \subset B^N$ and a one-to-one map $\varphi: V \to W$ such that $\mu^{(N)}(V) \to 1$ and $\liminf_{N \to \infty} N^{-1} \log|W| \geqq R > H$. Let ε, δ be numbers such that $0 < \varepsilon, \delta < 1$. Then for N sufficiently large, there exists a f.d. Rohlin N-set $F \subset A^\infty$, sets $S_1 \subset S_2 \subset V$, a map $\psi: W \to S_2$ and jointly ergodic processes U, X with respective state spaces A, B such that:*

(a) *the distribution of U is μ.*

(b) $N Pr[U^{2N} \in S_1 \times S_1, U \in F, T_A^N U \in F] > 1 - 2\varepsilon$,
$N Pr[U^{2N} \in S_2 \times S_2, U \in F, T_A^N U \in F] > 1 - \delta$.

(c) *With probability* 1, *if* $U \in F$ *and* $U^N \in S_2$ *then* $X^N \in W$, $\psi(X^N) = U^N$; *if* $U \in F$ *and* $U^N \in S_1$, *then* $X^N = \varphi(U^N)$.

(d) $H(X) \geqq H + \varepsilon(R-H) - \delta$.

Proof. Choose a positive $\varepsilon' < \varepsilon$ so that $\varepsilon'(R-H) > -\delta + \varepsilon(R-H)$. Find for each N a set $S_2 \subset V$ so that $\mu^{(N)}(S_2) \to 1$ and $\lim_{N \to \infty} N^{-1} \log|S_2| = H$. Find $S_1 \subset S_2$ so that $\mu^{(N)}(S_1) \to 1 - \varepsilon'$. Since $\liminf_{N \to \infty} N^{-1} \log|W - \varphi(S_1)| \geqq R$ and $\lim_{N \to \infty} N^{-1} \log|S_2 - S_1| = H$, there is a partition $\{W_u : u \in S_2 - S_1\}$ of $W - \varphi(S_1)$ such that

(e) $\liminf_{N \to \infty} N^{-1} [\min_{u \in S_2 - S_1} \log|W_u|] \geqq R - H$.

Define $\psi: W \to S_2$ to be the map such that

$$\begin{aligned} \psi &= \varphi^{-1} \quad \text{on } \varphi(S_1) \\ &= u \quad \text{on } W_u, \quad u \in S_2 - S_1. \end{aligned}$$

By a strong form of Rohlin's theorem [11, p. 22], for each N we can choose a f.d. Rohlin N-set $F \subset A^\infty$ such that

(f) $\liminf_{N \to \infty} N \mu[F \cap T_A^{-N} F \cap \{u : u^{2N} \in S_1 \times S_1\}] \geqq 1 - 2\varepsilon' > 1 - 2\varepsilon$.

(g) $\lim_{N \to \infty} N \mu] F \cap \{u : u^N \in S_2 - S_1\}] = \varepsilon'$.

(h) $\lim_{N \to \infty} N \mu[F \cap T_A^{-N} F \cap \{u : u^{2N} \in S_2 \times S_2\}] = 1$.

Fix $x^* \in B$. Let $[A, \tau, B]$ be a stationary channel such that for each $u \in A^\infty$,

(i) $\{\tilde{X}_i^N : T_A^i u \in F\}$ are indepencent under τ_u.

(j) $\tau_u[\tilde{X}_i^N = \varphi(u_i^N)] = 1, u_i^N \in S_1,\ T_A^i u \in F$.

(k) $\tau_u[\tilde{X}_i^N = x] = |W_{u_i^N}|^{-1}, x \in W_{u_i^N}, u_i^N \in S_2 - S_1,\ T_A^i u \in F$.

(l) $\tau_u[\tilde{X}_j = x^*] = 1, j \notin \bigcup_{T_A^i u \in F} [i, i+N-1]$.

Let U, X be processes with respective state spaces A, B, such that the distribution of U is μ, and the distribution of X conditioned on U is given by τ. Then U, X are jointly ergodic by [1], and (c) holds. Because of (c) and (h), $H(U|X) \to 0$ as $N \to \infty$. Because of (k), (g), and (e), $\liminf_{N \to \infty} H(X|U) \geqq \varepsilon'(R-H)$. Hence

$$\begin{aligned} \liminf_{N \to \infty} [H(X) - H(U)] &= \liminf_{N \to \infty} [H(X|U) - H(U|X)] \\ &\geqq \varepsilon'(R-H) > -\delta + \varepsilon(R-H), \end{aligned}$$

giving (d). Property (b) holds because of (f), (h).

5. Proof of Theorem 1

Lemma 11. *Let* $[B, v, C]$ *be stationary and let* $\tau \in \mathscr{P}_e(B)$. *Suppose* $\mathscr{N}$ *is a neighborhood of* τ *in* $\mathscr{P}_e(B)$ *such that for every positive entropy measure* λ *in* $\mathscr{N}$, Φ_v *is continuous at* λ *and* $[B, v, C]$ *is ergodic at* λ. *Then given* k *(a positive integer),* $\delta (\delta > 0)$, $R (0 < R < I(\tau v))$ *and* $\varepsilon (0 < \varepsilon < 1)$, *there exists* $\alpha (0 < \alpha < 1)$ *and for* n

sufficiently large a $(n, M, \varepsilon, \alpha)$ *sliding-block channel code* (W, G, g) *with* $M \geqq 2^{nR}$ *and every sequence in* W (k, δ) *typical of* τ.

Proof. Choose a stationary channel $[B, \sigma, C]$ such that $\mu\sigma = \mu\nu$, $\mu \in \mathscr{P}_e(B)$ and the map $\mu \to \mu\sigma$ from $\mathscr{P}_e'(B) \to \mathscr{P}(B, C)$ is continuous at any measure in $\mathscr{P}_e(B)$ if and only if Φ_ν is, where $\mathscr{P}_e'(B)$ is the set of all T_B-stationary and ergodic measures on B^∞, with the weak topology. Call this map Φ_σ'. There must be a neighborhood $\mathscr{N}'$ of τ in $\mathscr{P}_e'(B)$ such that for every positive entropy measure λ in $\mathscr{P}_e'(B)$, $[B, \sigma, C]$ is ergodic at λ and Φ_σ' is continuous at λ. By the proofs of Theorem 2 and Lemma 3 of [8], there exists α and for n sufficiently large a $(n, M, \varepsilon, \alpha)$ code (W, G, g) for the channel $[B, \sigma, C]$ with $M \geqq 2^{nR}$ and every sequence in $W(k, \delta)$ typical of τ. But a $(n, M, \varepsilon, \alpha)$ code for the channel $[B, \sigma, C]$ must be a $(n, M, \varepsilon, \alpha)$ code for the channel $[B, \nu, C]$, since $\mu\nu = \mu\sigma$, $\mu \in \mathscr{P}_e(B)$.

We now proceed with the proof of Theorem 1. Fix $[A, \mu], \tau, [B, \nu, C]$ given in the statement of Theorem 1. Fix $\beta > 0$. We will find a ν-invulnerable $[B, \lambda]$ isomorphic to $[A, \mu]$ for which $d_w(\lambda, \tau) < \beta$.

Fix neighborhoods $\mathscr{N}, \mathscr{N}'$ of τ in $\mathscr{P}_e(B)$ so small that

(a) $[B, \nu, C]$ is ergodic at every positive entropy measure in $\mathscr{N}$ and Φ_ν is continuous at every measure in $\mathscr{N}$.

(b) If $\{\lambda_n\} \subset \mathscr{N}'$, $\lambda \in \mathscr{P}_e(B)$, and $\lambda_n \to \lambda$, then $d_w(\lambda, \tau) < \beta$ and $\lambda \in \mathscr{N}$.

Construct on some probability space a process U with state space A and distribution μ, processes $\{X(i)\}_1^\infty$ with state space B jointly stationary with U, positive numbers $\{\varepsilon_i\}_1^\infty$, $\{\delta_i\}_1^\infty$, positive integers $\{m_i\}_1^\infty$, $\{N_i\}_1^\infty$, sets $\{S_1^{(i)}\}_1^\infty$, $\{S_2^{(i)}\}_1^\infty$, $\{F_i\}_1^\infty$, $\{W_i\}_1^\infty$, $\{G_i\}_1^\infty$, and functions $\{\varphi_i\}_1^\infty$, $\{\psi_i\}_1^\infty$, $\{g_i\}_1^\infty$ such that for each $i \geqq 1$:

(c) $U, X(i)$ are jointly ergodic and the distribution λ_i of $X(i)$ lies in $\mathscr{N}'$.

(d) $m_i > N_i$, $F_i \subset A^{2m_i+1}$ and $\{u \in A^\infty : u_{-m_i}^{2m_i+1}\}$ is a Rohlin N_i-set; $W_i \subset B^{N_i}$ and $\{x \in B^\infty : x^{2N_i} \in W_i \times W_i\}$ is a Rohlin N_i-set; $G_i \subset C^\infty$ is a f.d. Rohlin N_i-set.

(e) $S_1^{(i)} \subset S_2^{(i)} \subset A^{N_i}$, $N_i Pr[U_{-m_i}^{2m_i+1}, U_{N_i-m_i}^{2m_i+1} \in F_i, U^{2N_i} \in S_1^{(i)} \times S_1^{(i)}] > 1 - 2\varepsilon_i$,
and

$$N_i Pr[U_{-m_i}^{2m_i+1}, U_{N_i-m_i}^{2m_i+1} \in F_i, U^{2N_i} \in S_2^{(i)} \times S_2^{(i)}] > 1 - \delta_i.$$

(f) $\varphi_i: S_1^{(i)} \to W_i, \psi_i: W_i \to S_2^{(i)}$; with probability 1 if $U_{-m_i}^{2m_i+1} \in F_i$, $U^{N_i} \in S_2^{(i)}$ then $X(i)^{N_i} \in W_i$, $U^{N_i} = \psi_i(X(i)^{N_i})$, and if $U_{-m_i}^{2m_i+1} \in F_i$, $U^{N_i} \in S_1^{(i)}$, then $X(i)^{N_i} = \varphi_i(U^{N_i})$.

(g) $\delta_i < (56)^{-4}/2$, $\varepsilon_i < 1/2$, $\sum_i (\varepsilon_i + \delta_i^{\frac{1}{4}}) < \infty$.

(h) $H(X(i)) > H(\mu) + q_B(8\delta_i^{\frac{1}{4}}) + q_A(24\delta_i^{\frac{1}{4}})$.

(i) $g_i: C^\infty \to W_i$ is f.d. and for every $k \geqq i$,

$$N_i \lambda_k \nu[\bar{X}^{2N_i} \in W_i \times W_i, \bar{Y} \in G_i, \bar{X}^{N_i} = g_i(\bar{Y})] > 1 - \delta_i.$$

(j) If $u \in S_1^{(i+1)}$, for at least $N_i^{-1} N_{i+1}(1 - 2\varepsilon_i - 56\delta_i^{\frac{1}{4}})$ of the integers $j \in [m_i + 1, N_{i+1} - N_i - m_i]$ one has $u_{j-m_i}^{2m_i+1}$, $u_{j+N_i-m_i}^{2m_i+1} \in F_i$, $u_j^{2N_i} \in S_1^{(i)} \times S_1^{(i)}$, and $\varphi_{i+1}(u)_j^{N_i} = \varphi_i(u_j^{N_i})$.

(One begins the construction for $i = 1$ using Lemma 11 and Lemma 10. Having done the construction for $i = k$, say, one does it for $i = k + 1$ using Lemmas 8–10.) Define the sequence $\{\eta_i\}_1^\infty$ by $\eta_i = \sum_{k=i}^{\infty} (2\varepsilon_k + 56\delta_k^{\frac{1}{4}})$, $i \geqq 1$. From (j) we obtain

(k) If $k>i$, and $u\in S_1^{(k)}$, for at least $N_i^{-1}N_k(1-\eta_i)$ of the integers $j\in[m_i+1, N_k - N_i - m_i]$ one has $u_{j-m_i}^{2m_i+1}, u_{j+N_i-m_i}^{2m_i+1}\in F_i$, $u_j^{2N_i}\in S_1^{(i)}\times S_1^{(i)}$, and $\phi_k(u)_j^{N_i}=\varphi_i(u_j^{N_i})$.

For each $i\geqq 1$, let $\bar{\varphi}_i\colon A^\infty\to B^\infty, \bar{\psi}_i\colon B^\infty\to A^\infty, \bar{g}_i\colon C^\infty\to B^\infty$ be f.d. stationary codes such that

(l) $\bar{\varphi}_i(u)_j^{N_i}=\varphi_i(u_j^{N_i}),\ u_{j-m_i}^{2m_i+1}\in F_i,\ u_j^{N_i}\in S_1^{(i)}$.
$\bar{\psi}_i(x)_j^{N_i}=\psi_i(x_j^{N_i}),\ x_j^{2N_i}\in W_i\times W_i$.
$\bar{g}_i(y)_j^{N_i}=g_i(T_C^j y),\ T_C^j y\in G_i$.

From (k), (f), (l), it follows that for $k>i\geqq 1$, each of the three quantities $Pr[X(k)_0\neq X(i)_0]$, $Pr[\bar{\psi}_i(X(k))_0\neq U_0]$, $Pr[\bar{\varphi}_i(U)_0\neq X(k)_0]$ is no greater than $\eta_i + 1 - N_k Pr[U_{-m_k}^{2m_k+1}\in F_k,\ U^{N_k}\in S_1^{(k)}]$. This implies there must be a process X with state space B, such that $\lim_{k\to\infty} Pr[X(k)_0\neq X_0]=0$ and hence for each $i\geqq 1$,

(m) $Pr[\bar{\psi}_i(X)_0\neq U_0], Pr[\bar{\varphi}_i(U)_0\neq X_0]\leqq\eta_i$.

Letting λ be the distribution of X, we see from (m) that $[B,\lambda]$ and $[A,\mu]$ are isomorphic. Since $\lambda_i\to\lambda$ weakly, from (c), (b) we have $d_w(\lambda,\tau)<\beta$, $\lambda\in\mathcal{N}$. From (i) and the definition of $\bar{g}_i$, if $k>i\geqq 1$,

$$\lambda_k\, v[\bar{X}_0=\bar{g}_i(\bar{Y})_0]>1-\delta_i.$$

Letting $k\to\infty$, since Φ_v is continuous at λ, we see that

$$\lambda\, v[\bar{X}_0=\bar{g}_i(\bar{Y})_0]\geqq 1-\delta_i,\qquad i\geqq 1.$$

Hence $[B,\lambda]$ is v-invulnerable.

References

1. Adler, R.L.: Ergodic and Mixing Properties of Infinite Memory Channels. Proc. Amer. Math. Soc., **12**, 924–930 (1961)
2. Kieffer, J.C.: Perfect Transmission Over a Discrete Memoryless Channel Requires Infinite Expected Coding Time. [To appear in the Journal of Combinatorics, Information and System Sciences]
3. Gray, R.M., Ornstein, D.S., Dobrushin, R.L.: Block Synchronization, Sliding-Block Coding, Invulnerable Sources, and Zero Error Codes for Discrete Noisy Channels. Ann. Probab. **8**, 639–674 (1980)
4. Lovasz, L.: On the Shannon Capacity of a Graph. IEEE Trans. Inform. Theory, **25**, 1–7 (1979)
5. Kieffer, J.C.: On sliding block coding for transmission of a source over a stationary nonanticipatory channel. Inform. and Control, **35**, 1–19 (1977)
6. Kieffer, J.C.: On the transmission of Bernoulli sources over stationary channels. Ann. Probab. **8**, 942–961 (1980)
7. Kieffer, J.C.: Block Coding for a Stationary Channel Satisfying a Weak Continuity Condition. [Submitted for publication]
8. Kieffer, J.C.: Sliding-Block Coding for Weakly Continuous Channels. [Submitted for publication]
9. Shannon, C.E.: The Zero Error Capacity of a noisy channel. Trans. Inform. Theory, **2**, 8–19 (1956)
10. Ornstein, D.S.: Bernoulli shifts with the same entropy are isomorphic. Adv. in Math., **4**, 337–352 (1970)
11. Shields, P.C.: The theory of Bernoulli shifts. Chicago: Univ. of Chicago 1973

Received June 14, 1980